KeyMap: A catalog of keys and mappings for linking Knowledge Graphs in the Web

Thibaut Soulard – Fatiha Saïs – Joe Raad – Gianluca Quercini

Atelier ROCED – PFIA'22

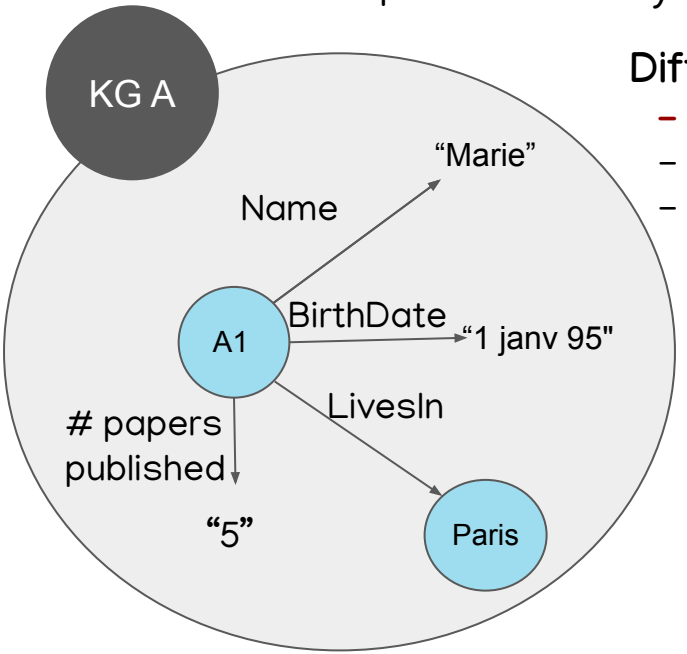
université PARIS-SACLAY

LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

cnrs

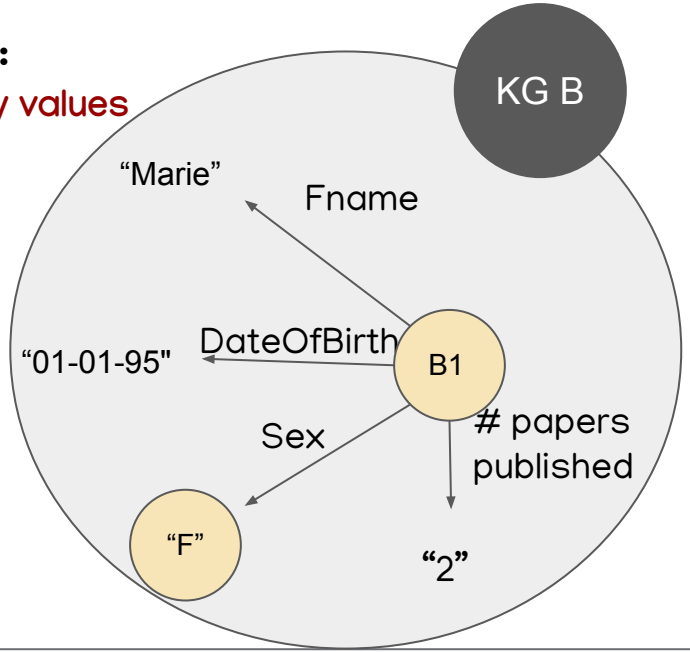
Entity linking : Motivations

An explainable Entity Linking method between N heterogeneous graphs.



Different approaches of linking:

- Similarity between all property values
- Graph embeddings
- Use of keys



Entity linking : Motivations

An explainable Entity Linking method between N heterogeneous graphs.

Different approaches of linking:

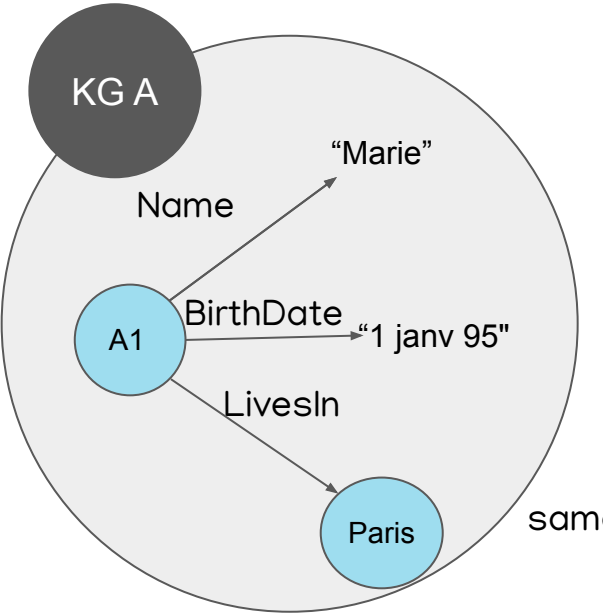
- Similarity between all property values
- Graph embeddings
- Use of keys

Graph Embedding can be used through link prediction, but requires seeds of already linked entities while not being easily explainable.

- *A Survey on Knowledge Graph Embedding: Approaches, Applications and Benchmarks*
Yuanfei Dai, Shiping Wang, Neal N. Xiong and Wenzhong Guo
- *A Benchmarking Study of Embedding-based Entity Alignment for Knowledge Graphs*
Zequn Sun, Qingheng Zhang, Wei Hu, Chengming Wang, Muhao Chen, Farahnaz Akrami, Chengkai Li

Entity linking : Motivations

An explainable Entity Linking method between N heterogeneous graphs.



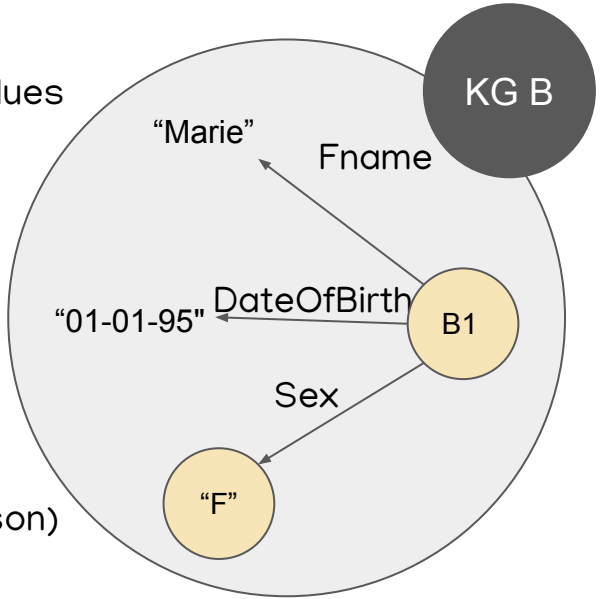
Different approaches of linking:

- Similarity between all properties values
- Graph embedding
- Use of keys

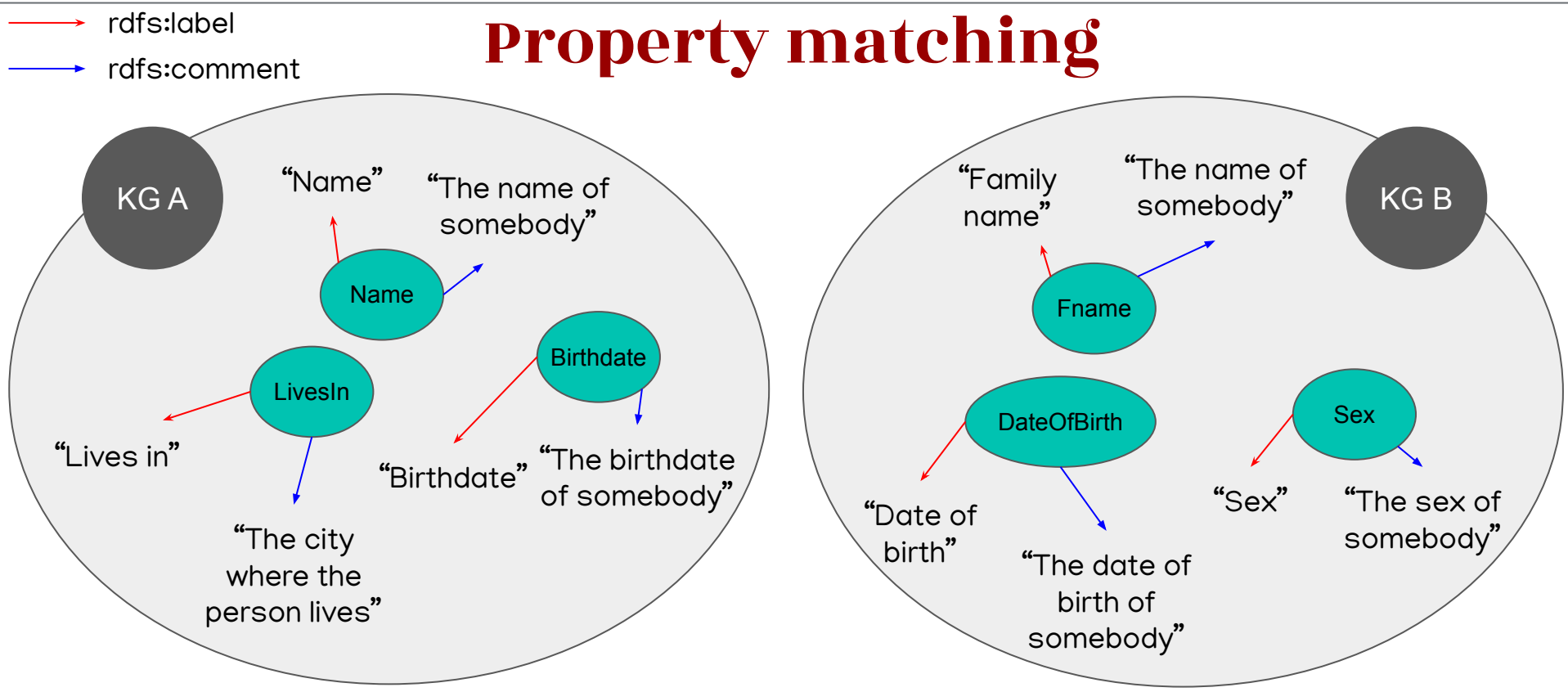
<u>Example of a key:</u>

K = Person:(Name, BirthDate)

same(Name) ∧ same(BirthDate) ⇒ same(Person)



Property matching



Property matching

KG A

	label	description	label doc	description doc
Name	name	The name of somebody	(name)	(name, somebody)
LivesIn	Lives in	The city where the person lives	(lives)	(city, person, lives)
Birthdate	Birth date	The birthdate of somebody	(birth, date)	(birthdate, somebody)

KG B

	label	description	label doc	description doc
fname	family name	The name of somebody	(family, name)	(name, somebody)
Sex	sex	The sex of somebody	(sex)	(sex, somebody)
DateOfBirth	Date of birth	The date of birth of somebody	(date, birth)	(date, birth, somebody)

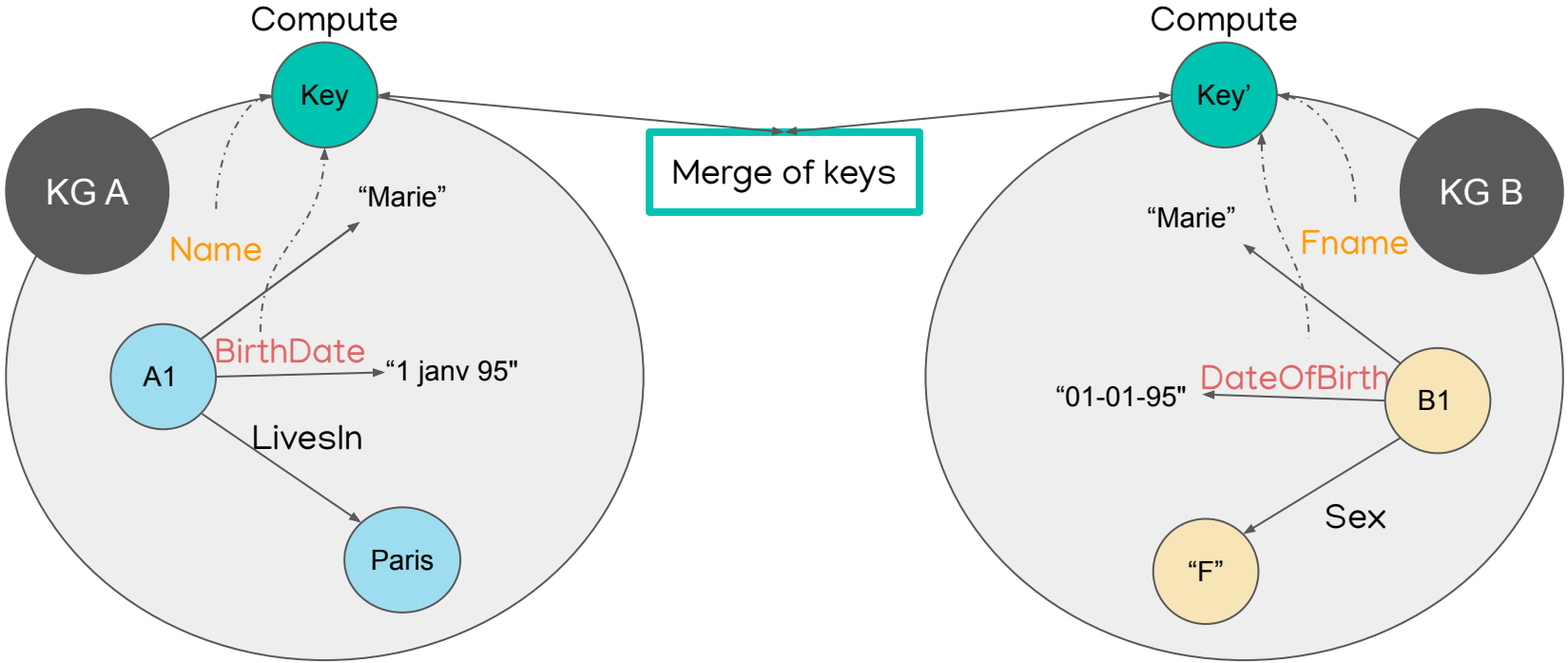
Property matching

Global Threshold

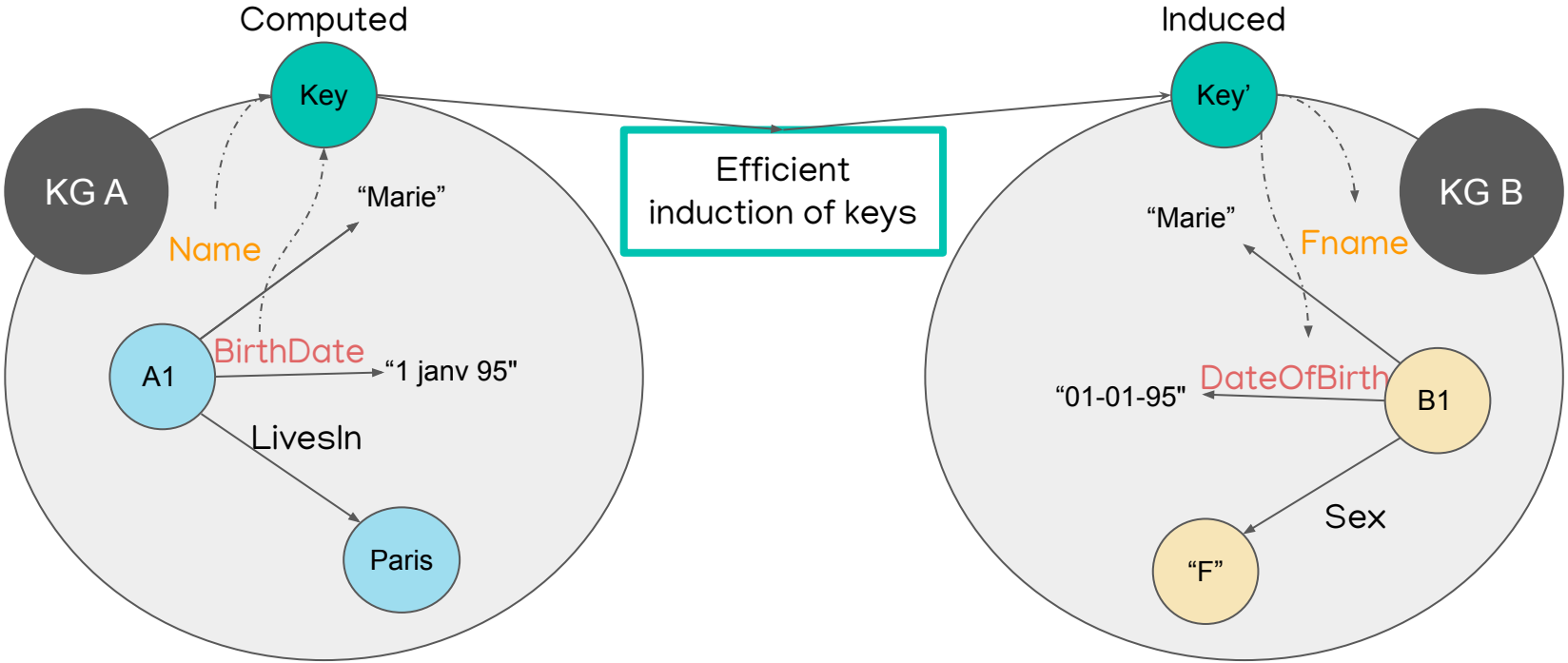
Relation A	Relation B	Similarity
Name	fname	0.90
Birthdate	DateOfBirth	0.86
Birthdate	fname	0.62
Name	DateOfBirth	0.58
LivesIn	fname	0.56
LivesIn	DateOfBirth	0.48
Birthdate	Sex	0.45
LivesIn	Sex	0.42
Name	Sex	0.40

Property matching

Use of keys for entity linking between two graphs



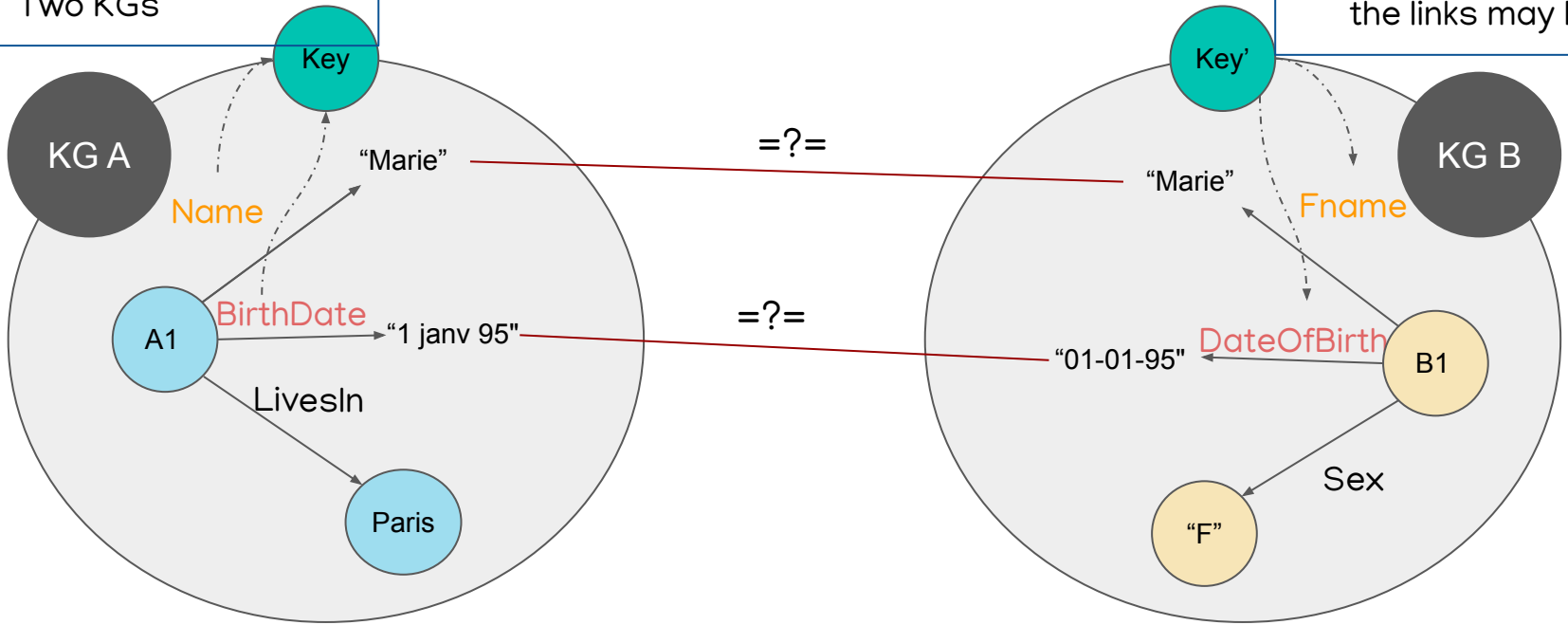
Use of keys for entity linking between two graphs



Key-based entity linking in KGs

Inputs:
- Property mappings
- Computed key
- Two KGs

Outputs :
- A set of identity links
- The keys that inferred the links may be kept.



How to generalise the approach to n graphs at Web scale ?

The approach:

- Build a catalog of keys and property mappings of n−1 graphs
- Exploit the catalog to find links between entities of a new graph G_n and the graphs G1, G2, … G_(n−1)

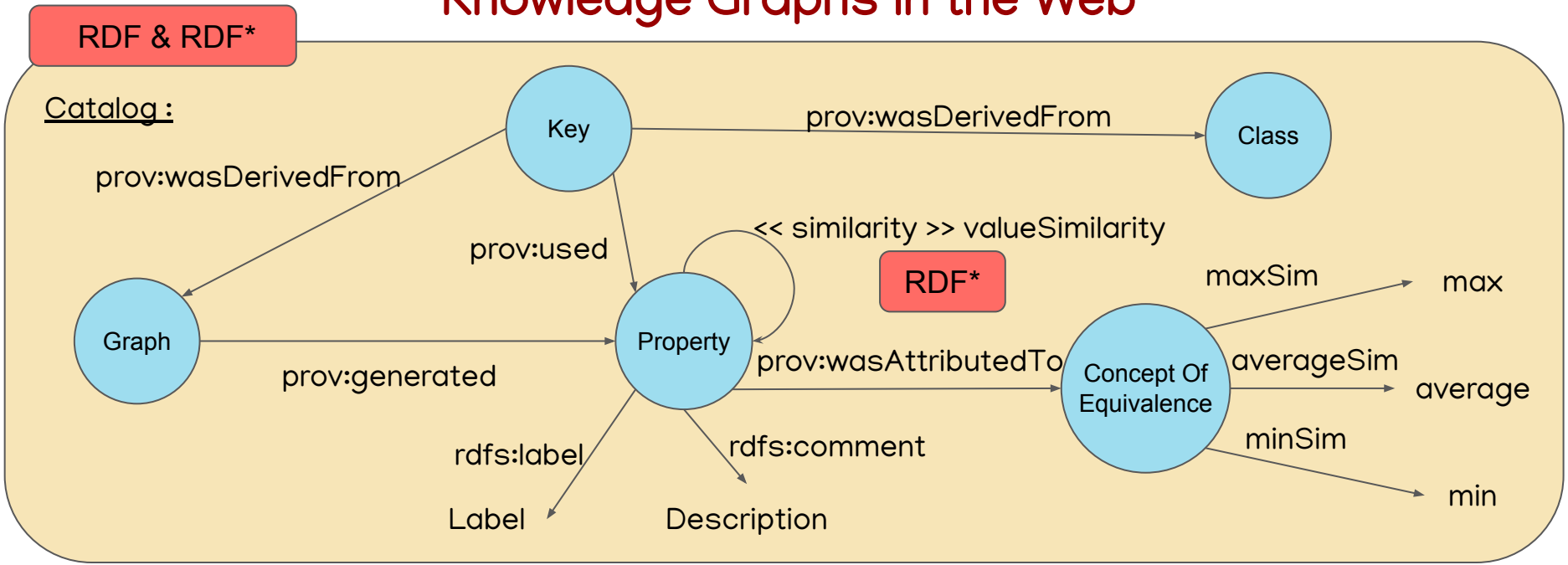
KeyMap: A catalog of keys and mappings for linking Knowledge Graphs in the Web

A graph that represents:

- The **description of properties** of n KGs
- The **result of property matching** of n KGs organised in equivalence classes while keeping the similarity scores
- The **result of key discovery** in n KGs
- Keep track of the **provenance information** for both keys and properties

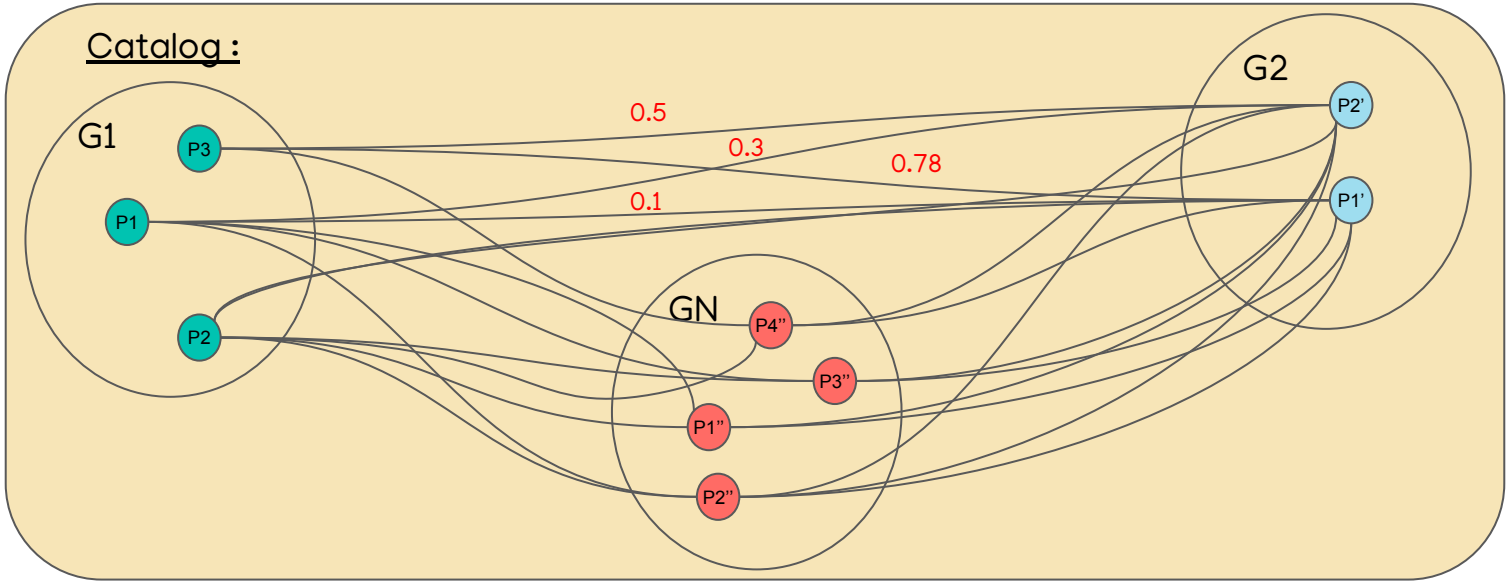
KeyMap catalog can be used for efficient linking of entities of a **new Knowledge Graph** with entities of **n Knowledge Graphs** that are already represented in the catalog

KeyMap: A catalog of keys and mappings for linking Knowledge Graphs in the Web



KeyMap: A catalog of keys and mappings for linking Knowledge Graphs in the Web

Inter-graph property mappings with their corresponding similarity scores (thanks to RDF*)

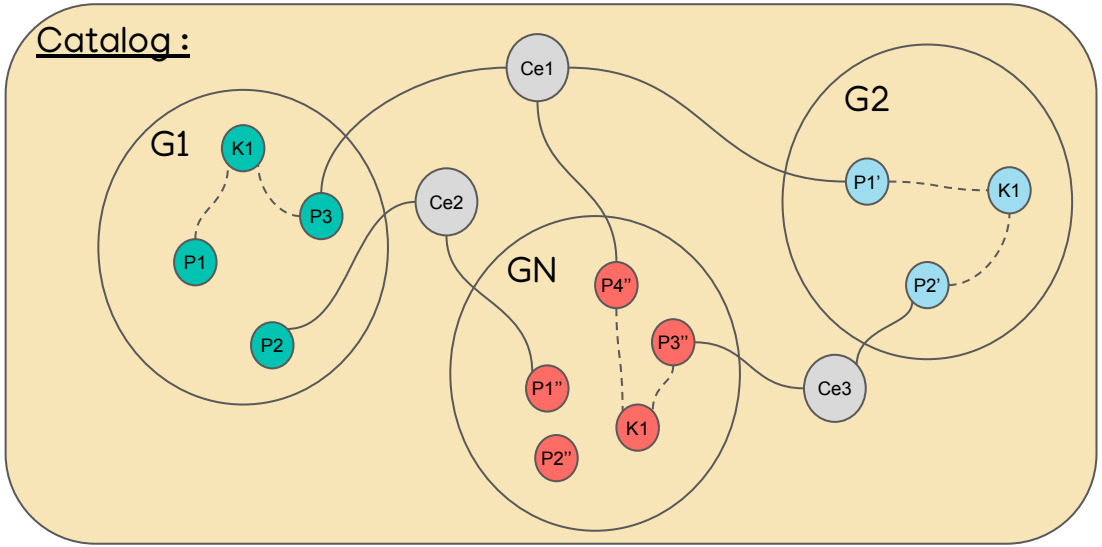


Scenario 1 - Entity linking without addition to the catalog

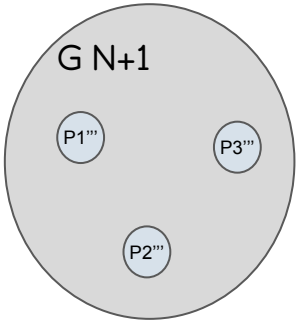
Only links with 1–n knowledges graphs.

The user wants to link his/her knowledge graph g to n–1 graphs described in the catalog, while keeping his/her knowledge graph private due to privacy concerns on the data.

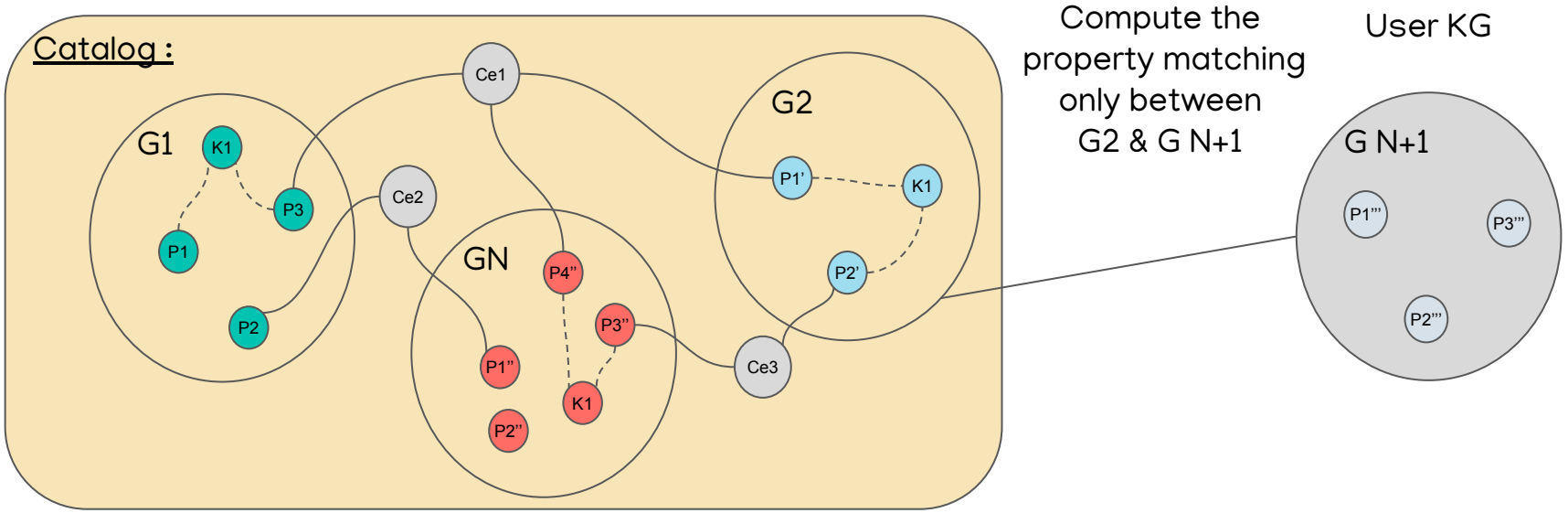
Scenario 1 - Entity linking without addition to the catalog



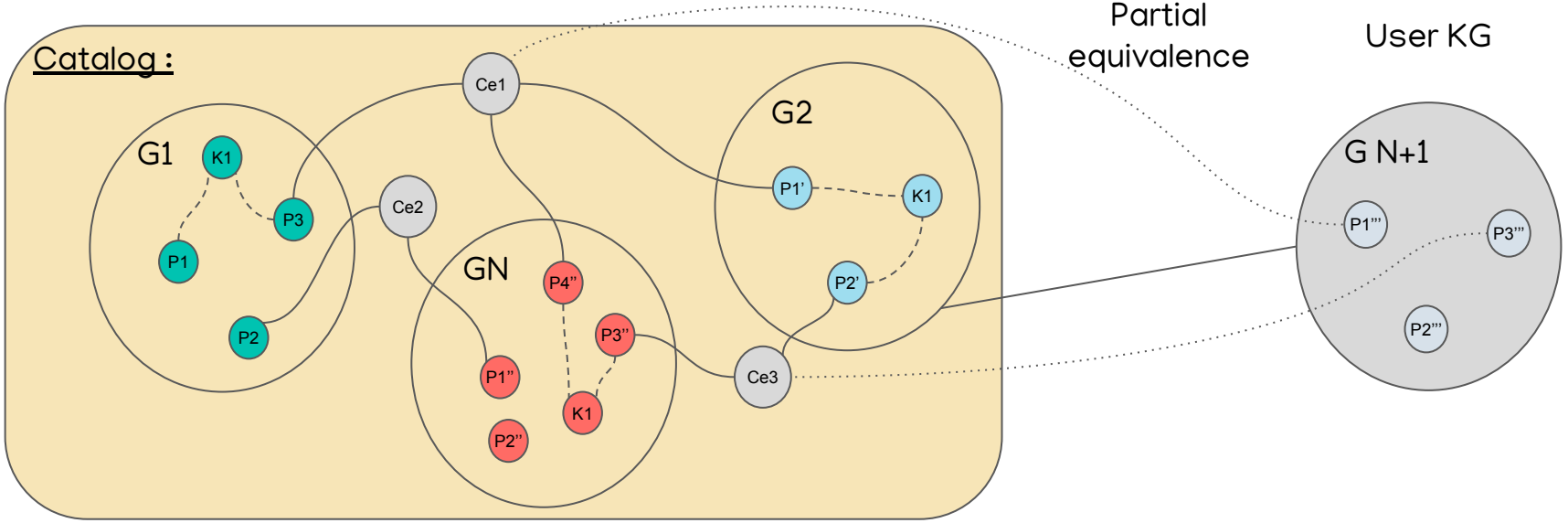
User KG



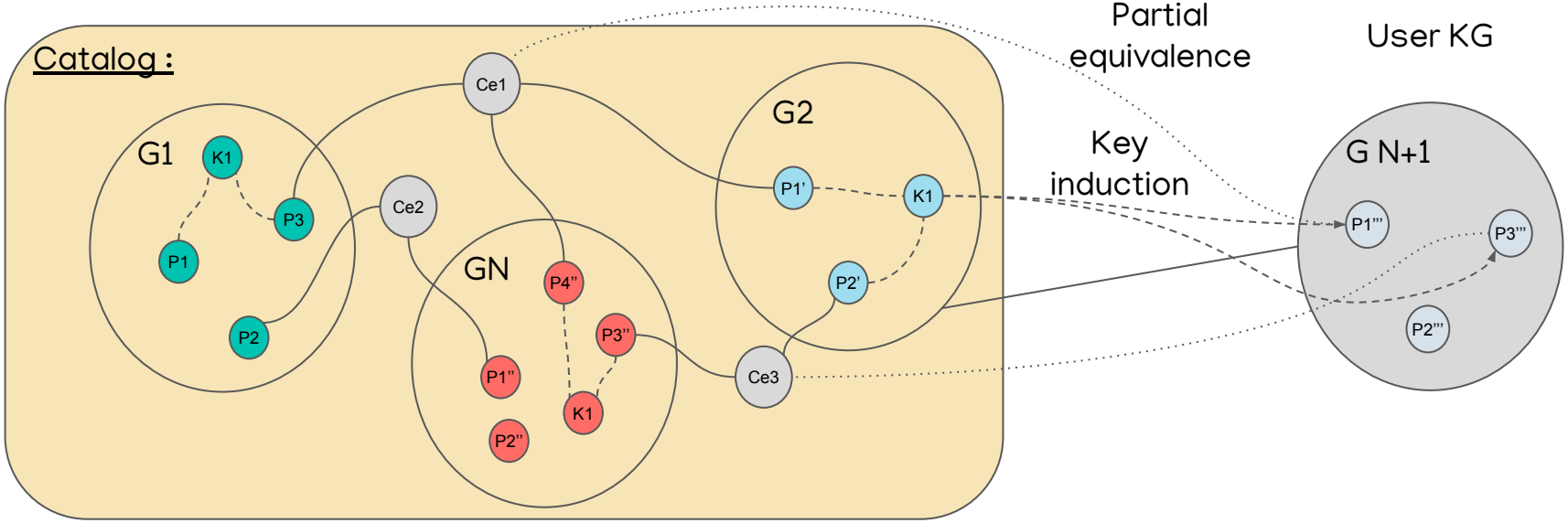
Scenario 1 - Entity linking without addition to the catalog



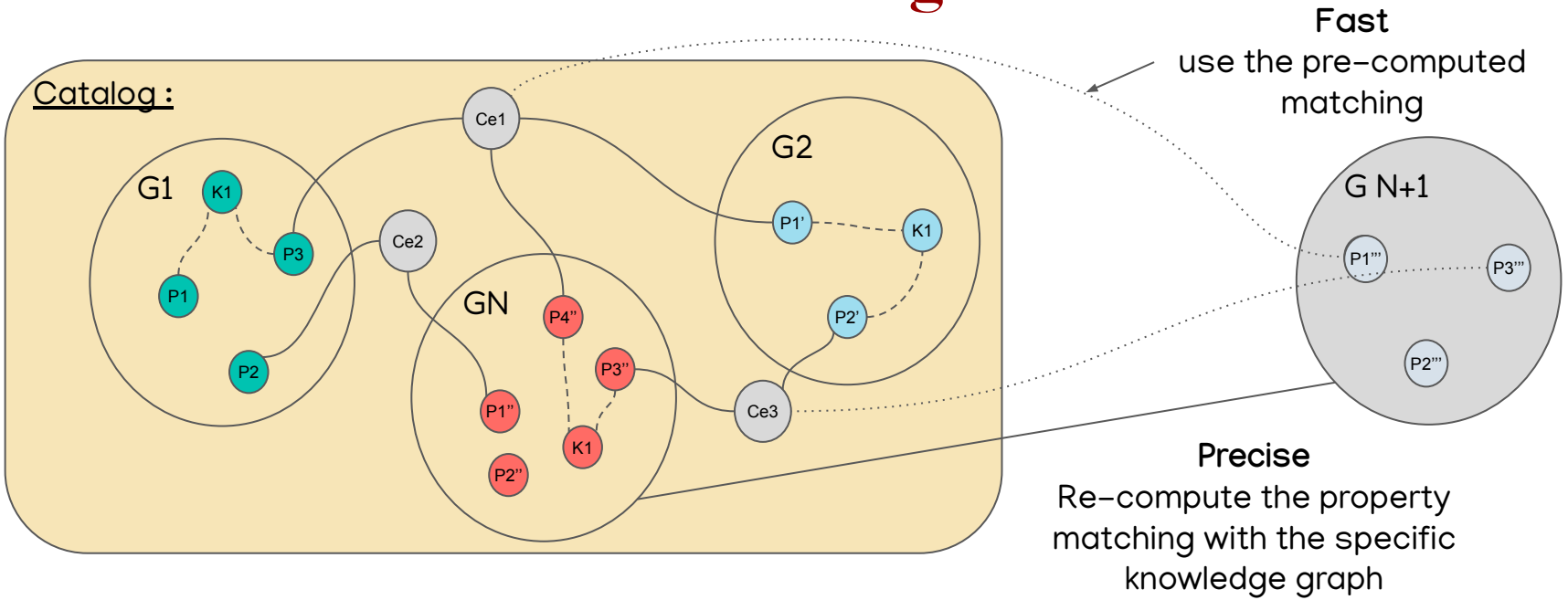
Scenario 1 - Entity linking without addition to the catalog



Scenario 1 - Entity linking without addition to the catalog



Scenario 1 - Entity linking without addition to the catalog

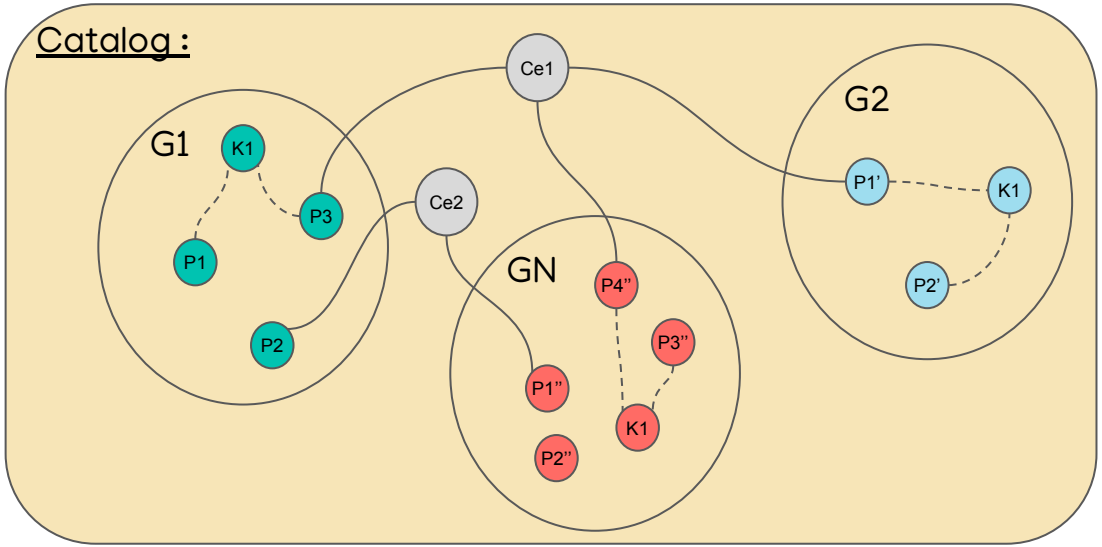


Scenario 2 - Incremental enrichment of the catalog

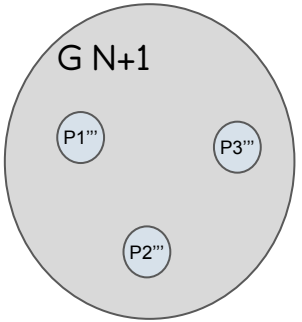
Integrate a new knowledge graph to the open catalog.

The user wants to link his/her graph g to n–1 graphs described in the catalog while enriching the catalog with the computed property mappings and keys to allow other users to use them

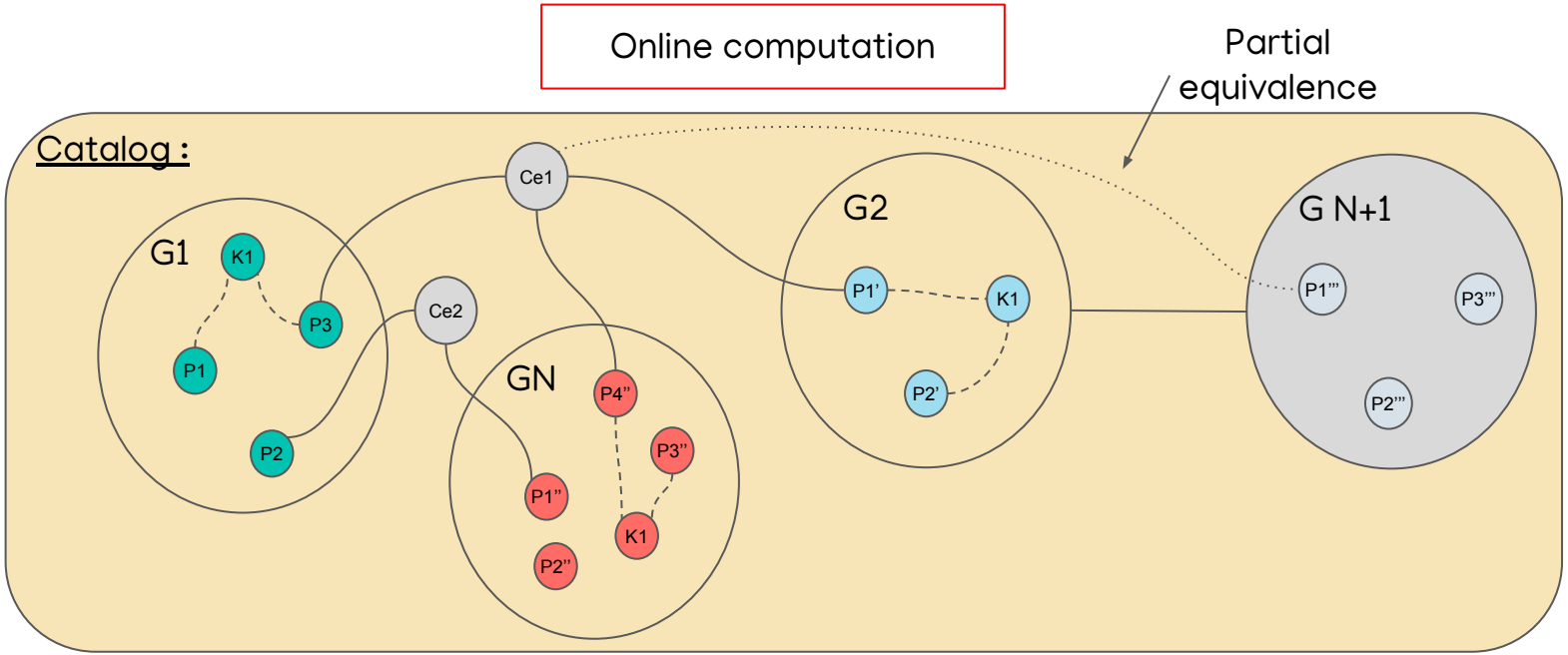
Scenario 2 - Incremental enrichment of the catalog



User KG

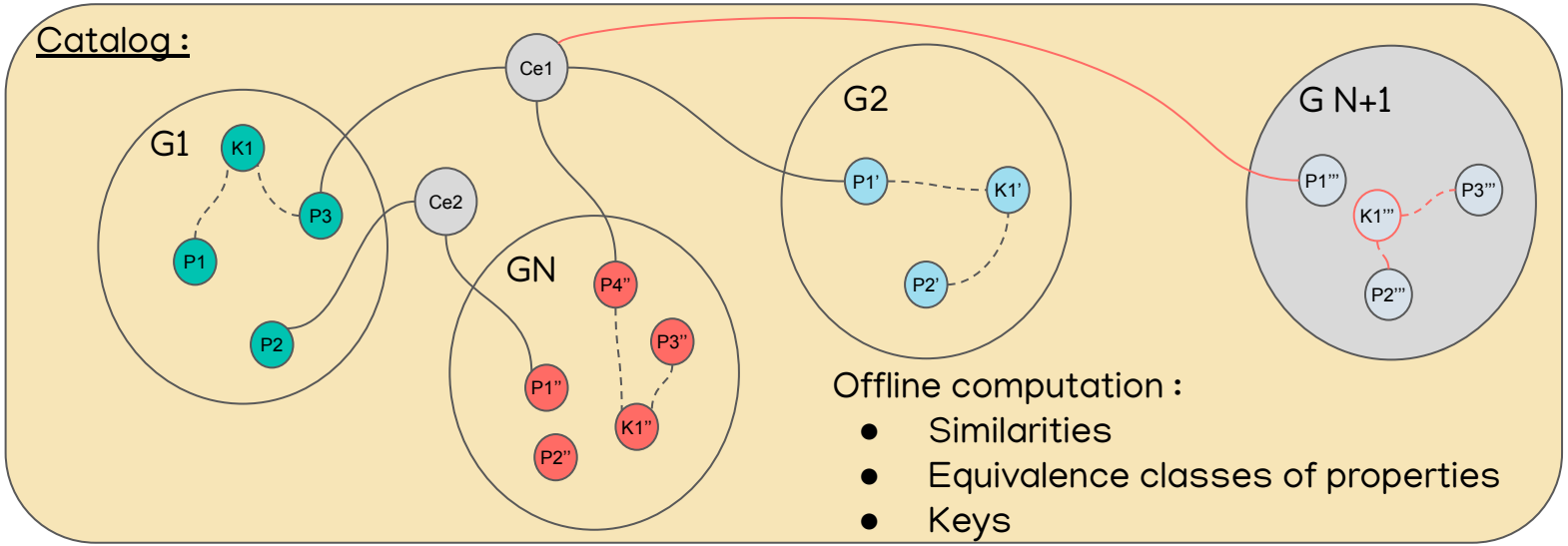


Scenario 2 - Incremental enrichment of the catalog



Scenario 2 - Incremental enrichment of the catalog

Offline computation



Conclusion

- A new efficient property matching
- A catalog to allow linking n graphs together
- A new key-based entity linking method that exploits the catalog

Future work

- Define a framework when knowledge graphs are updated
  - New/Deleted relations
  - Updated Name and/or Description

- Enrich the catalog with adapted similarity measures (e.g. Jaccard, Levenstein, ...)

- Explainability : represent identity links by attaching the keys that allowed their inference (possible with RDF *)

KeyMap: A catalog of keys and mappings for linking Knowledge Graphs in the Web

Thibaut Soulard – Fatiha Saïs – Joe Raad – Gianluca Quercini

Atelier ROCED – PFIA'22

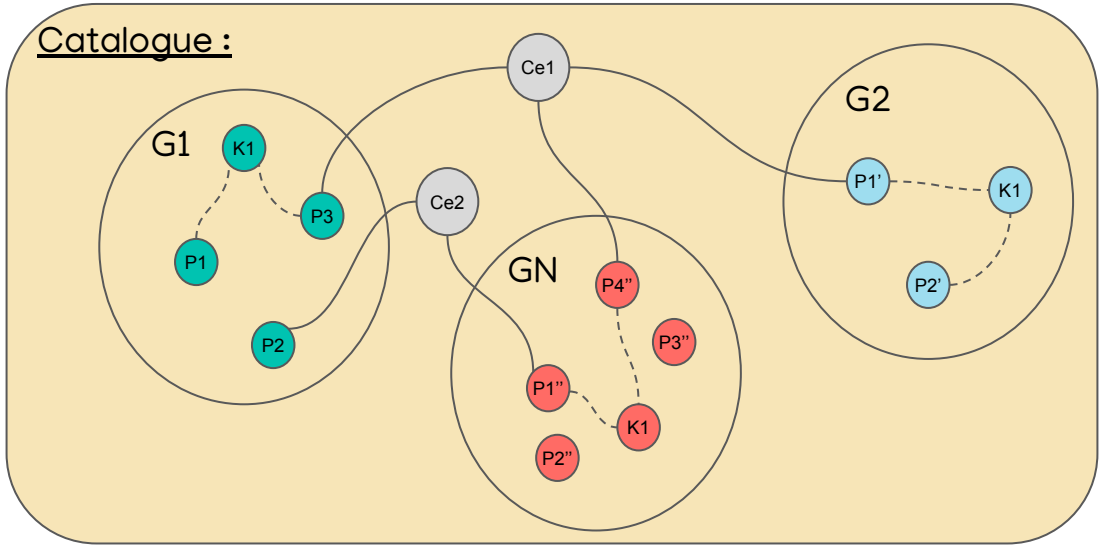
université PARIS-SACLAY

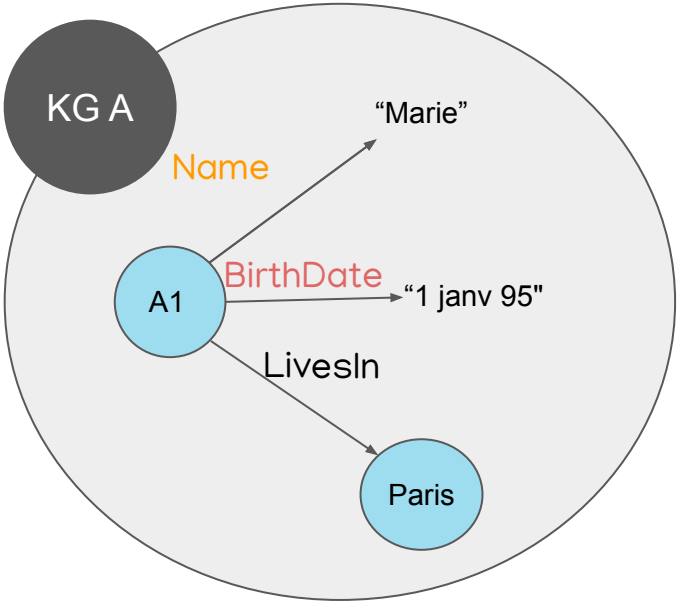
LISN LABORATOIRE INTERDISCIPLINAIRE DES SCIENCES DU NUMÉRIQUE

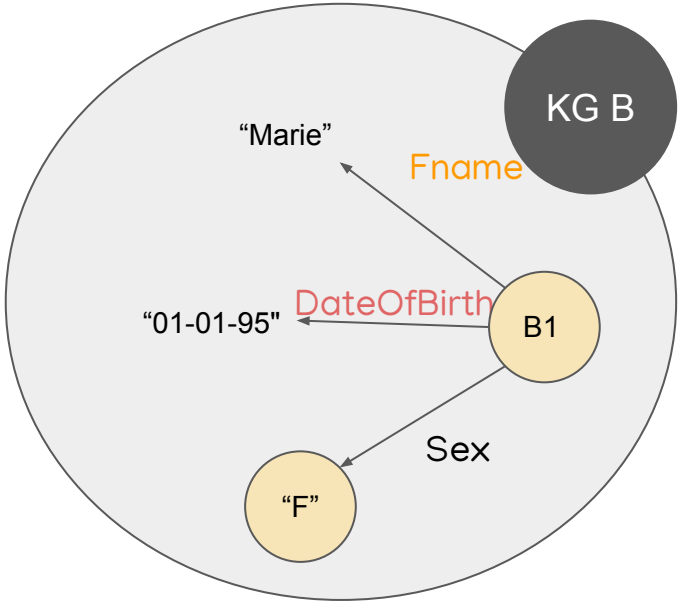
cnrs

Example


KG A
"Marie"
Name
A1
BirthDate
"1 janv 95"
LivesIn
Paris
KG B
"Marie"
Fname
"01-01-95"
DateOfBirth
B1
Sex
"F"


Catalogue :
Ce1
G1
K1
P3
P1
P2
Ce2
GN
P4''
P3''
P1''
K1
P2''
G2
P1'
K1
P2'


KG A
“Marie”
Name
A1
BirthDate
“1 janv 95"
LivesIn
Paris


KG B
“Marie”
Fname
“01-01-95"
DateOfBirth
B1
Sex
“F”

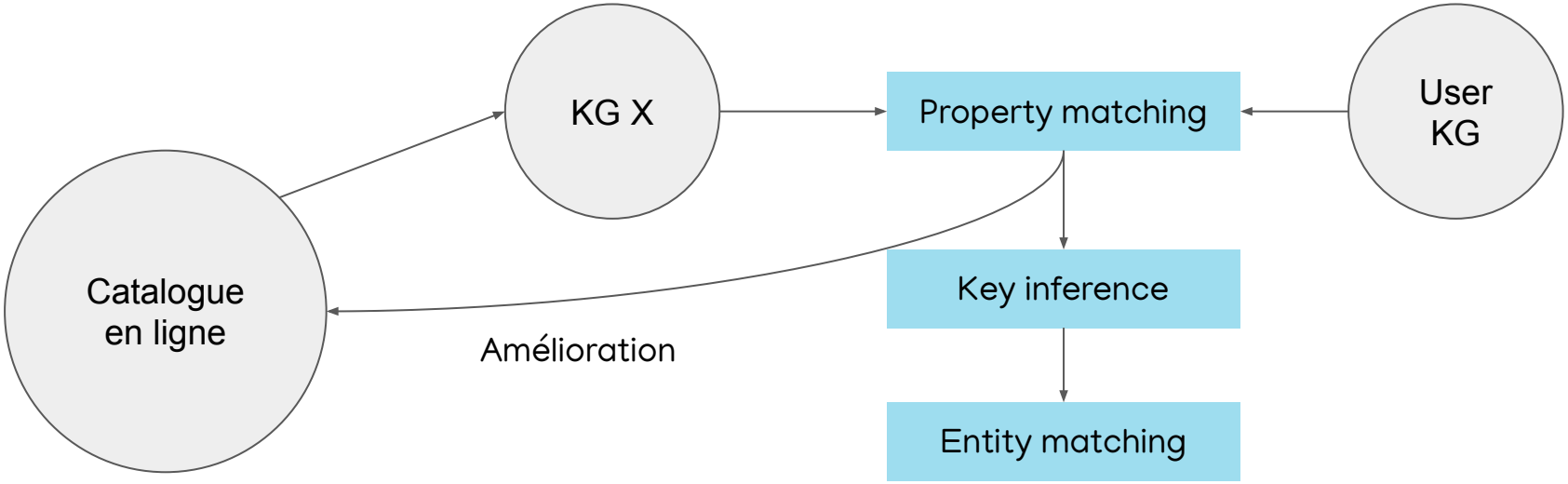
Thanks for your attention

Create an efficient and explainable catalog that allows user to link their knowledge graphs with N graphs.

NLP + keys

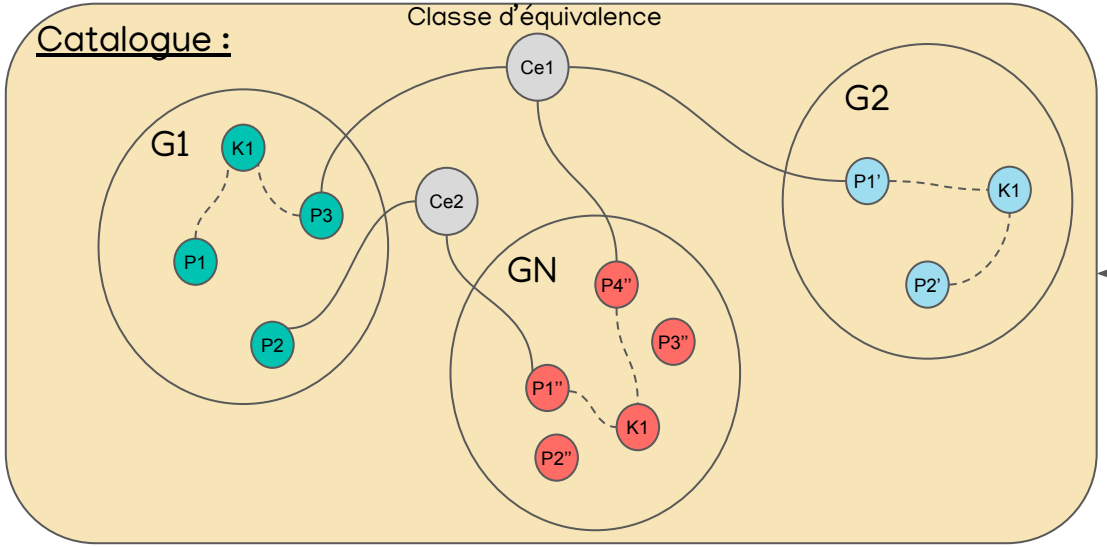
Available catalog to use and complete.

Résumé


KG X
Property matching
User KG
Catalogue en ligne
Amélioration
Key inference
Entity matching

Stage de Master 2 - Magistère 3

Liage d'entités dans des graphes de connaissances hétérogènes à l'échelle du web.



Alignement de schémas hétérogènes :

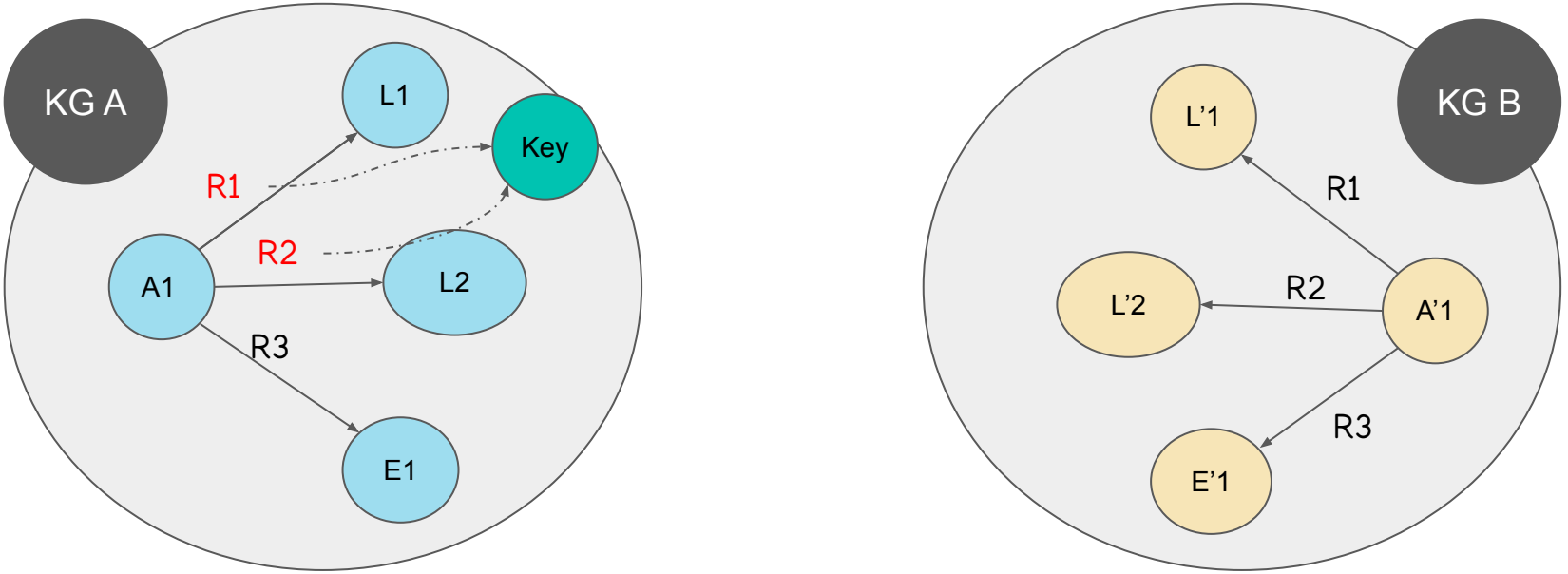
- Vectoriser "Nom" & "Description" des propriétés
- Similarité sur les vecteurs

User KG

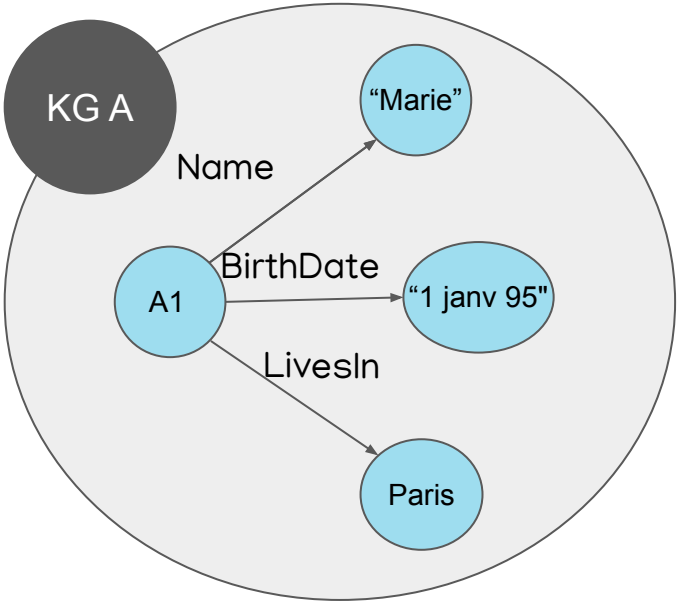
Découverte de clés pour le liage d'entités

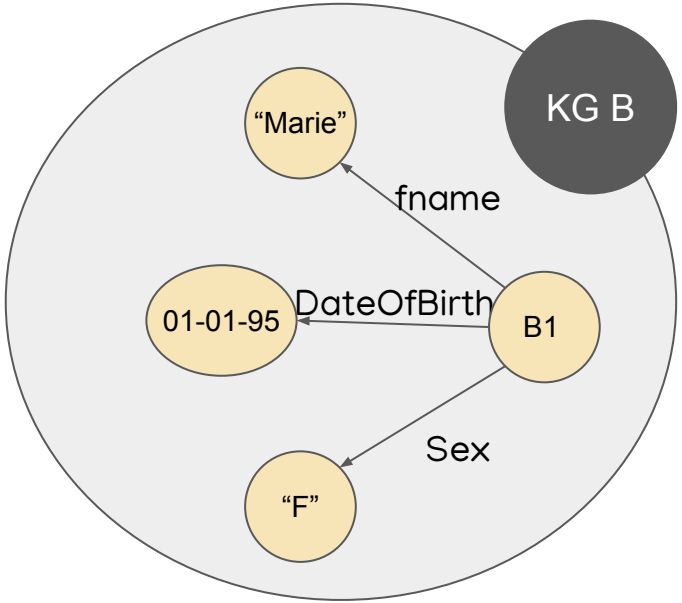
Test et évaluation en cours sur : wikidata, yago, dbpedia et schema.org

But
Property Mapping

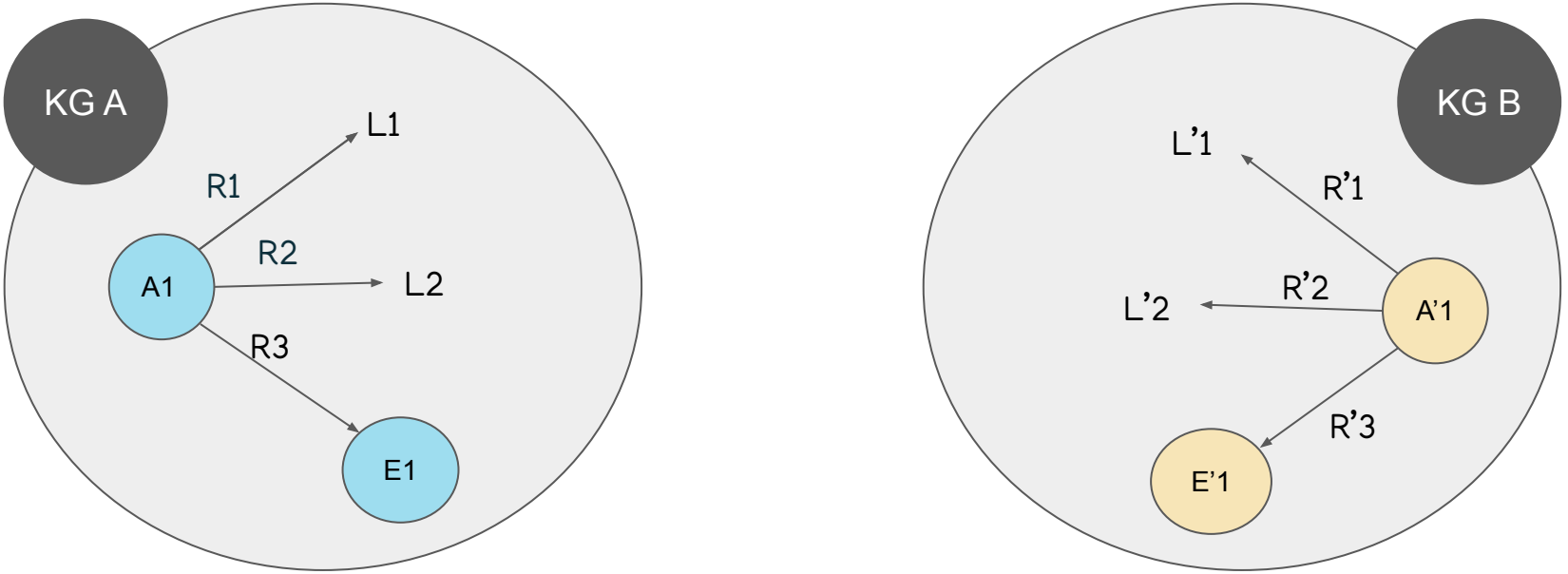


Problématique



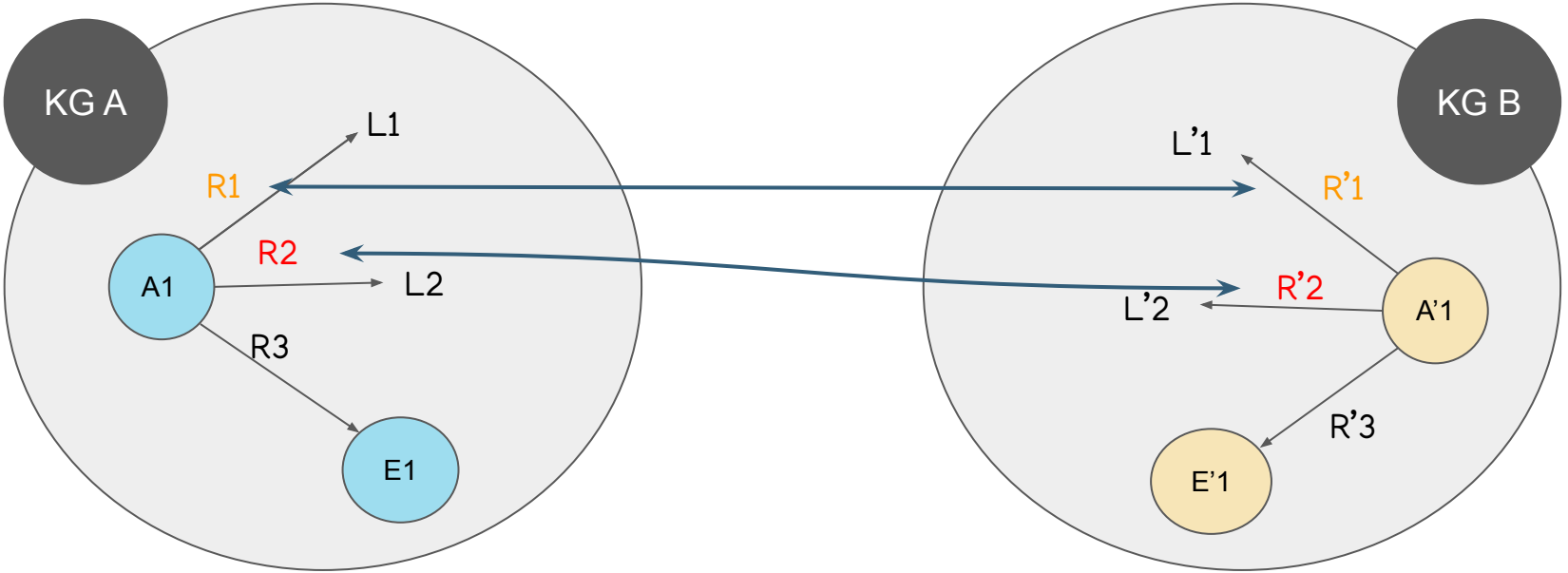


Idée générale
Heterogeneous graphs

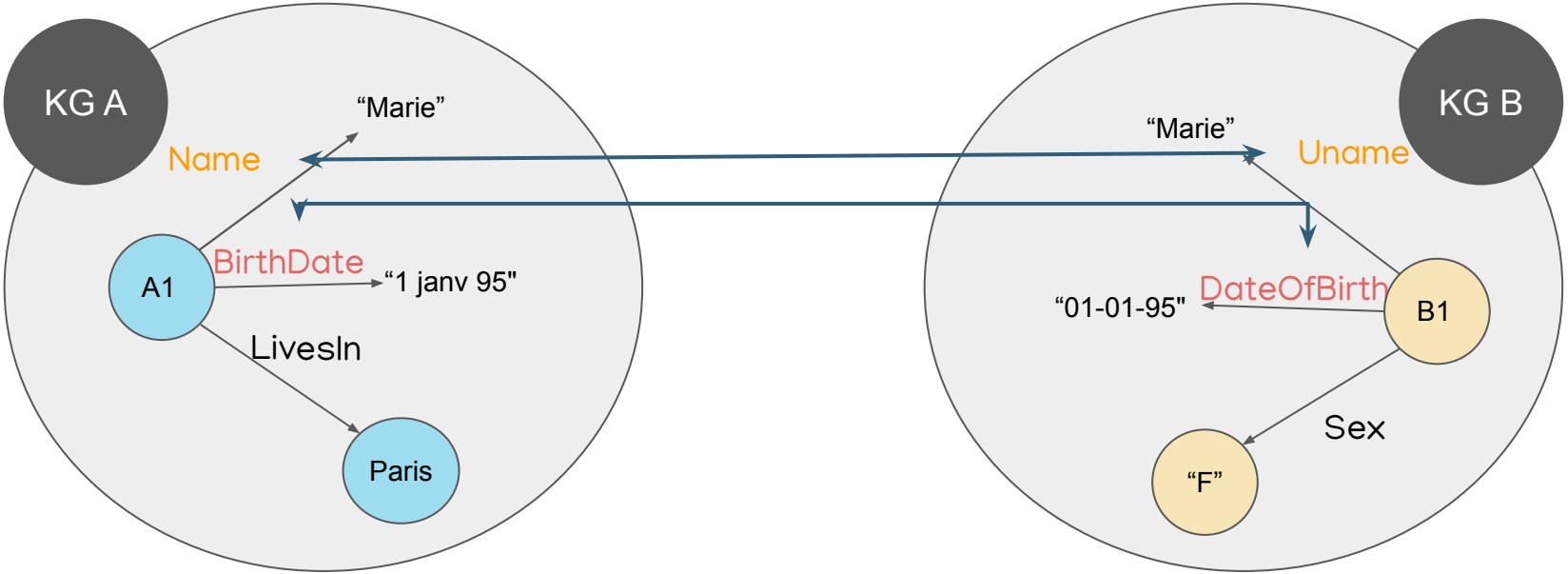


Idée générale
Heterogeneous graphs

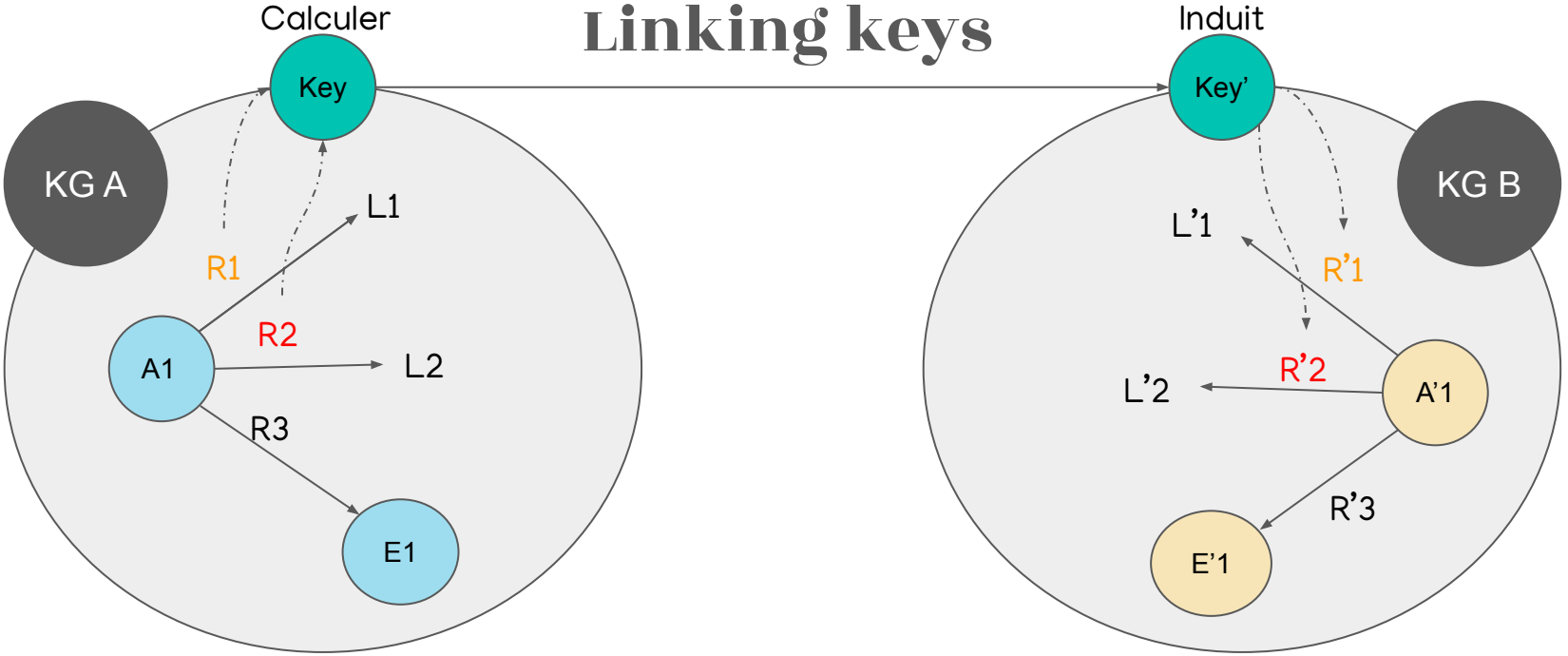
Idée générale
Property matching



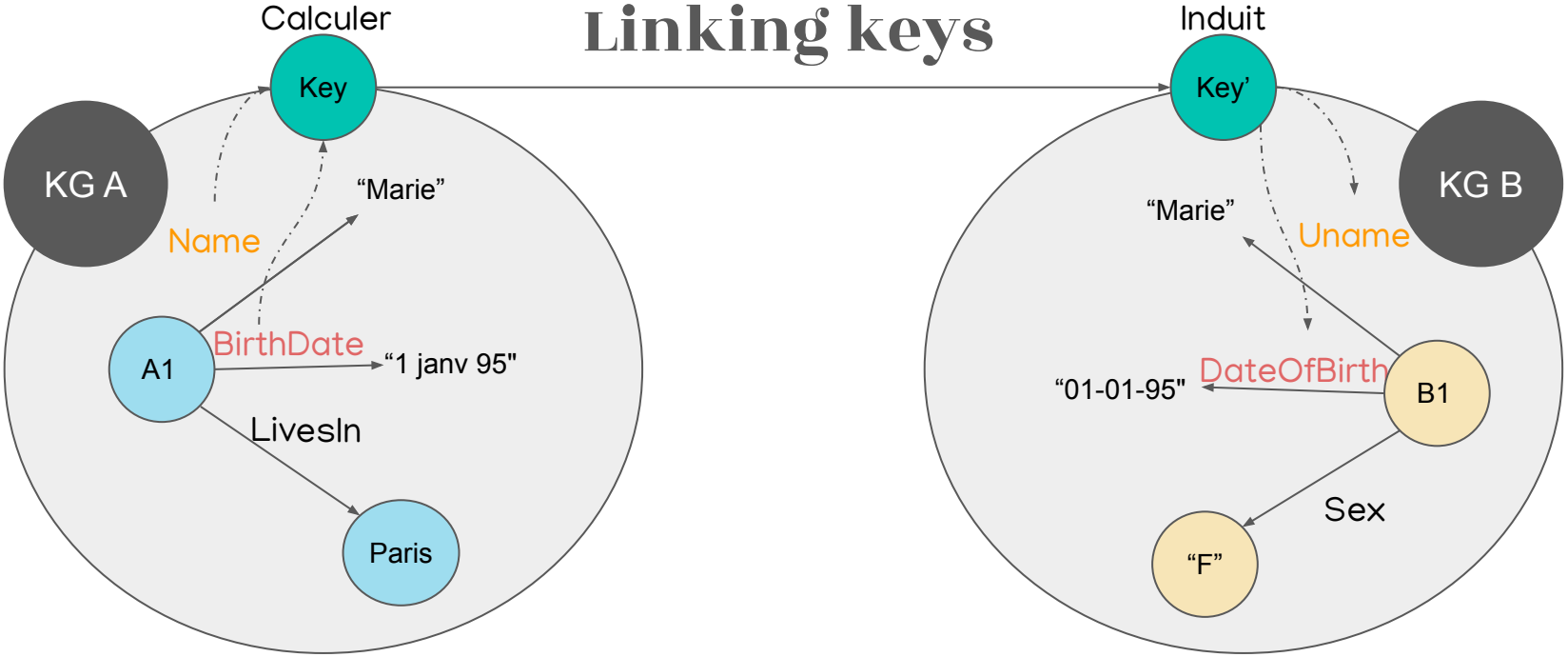
Idée générale
Property matching



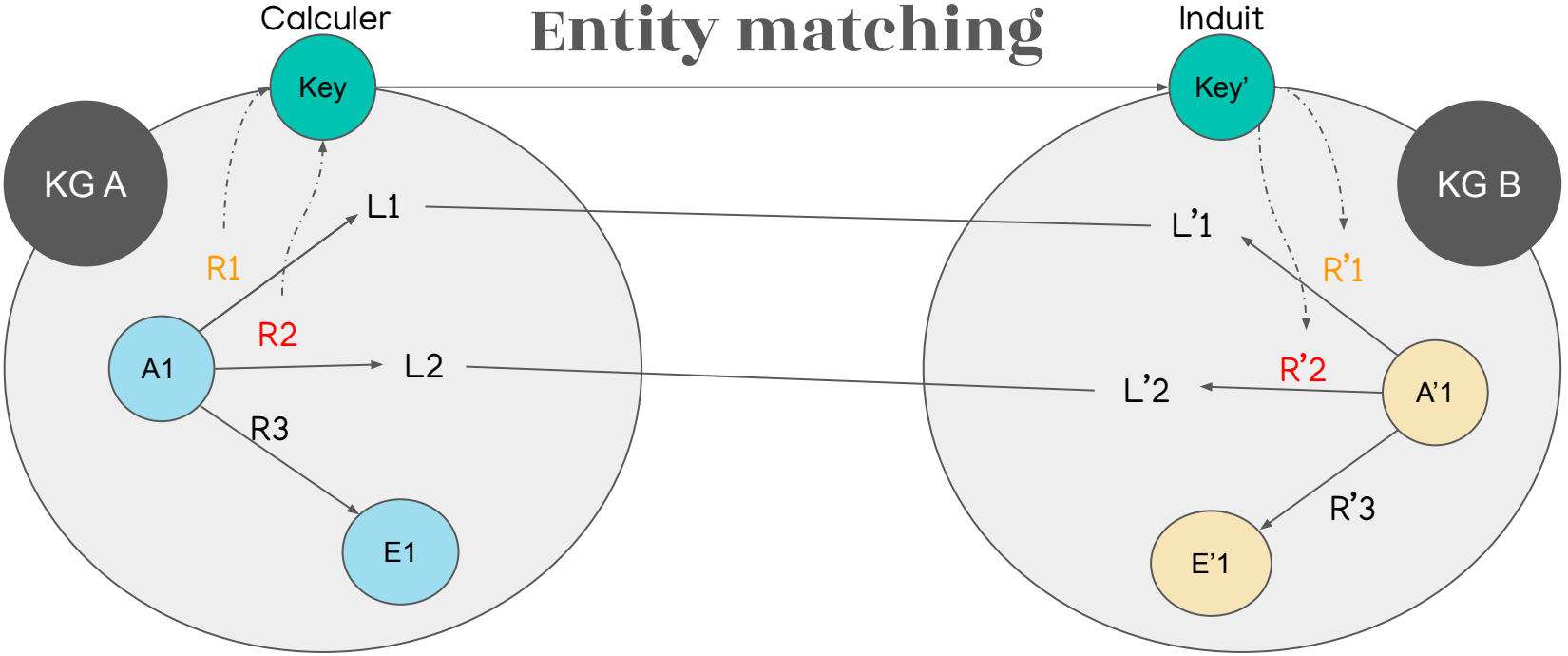
Idée générale
Linking keys



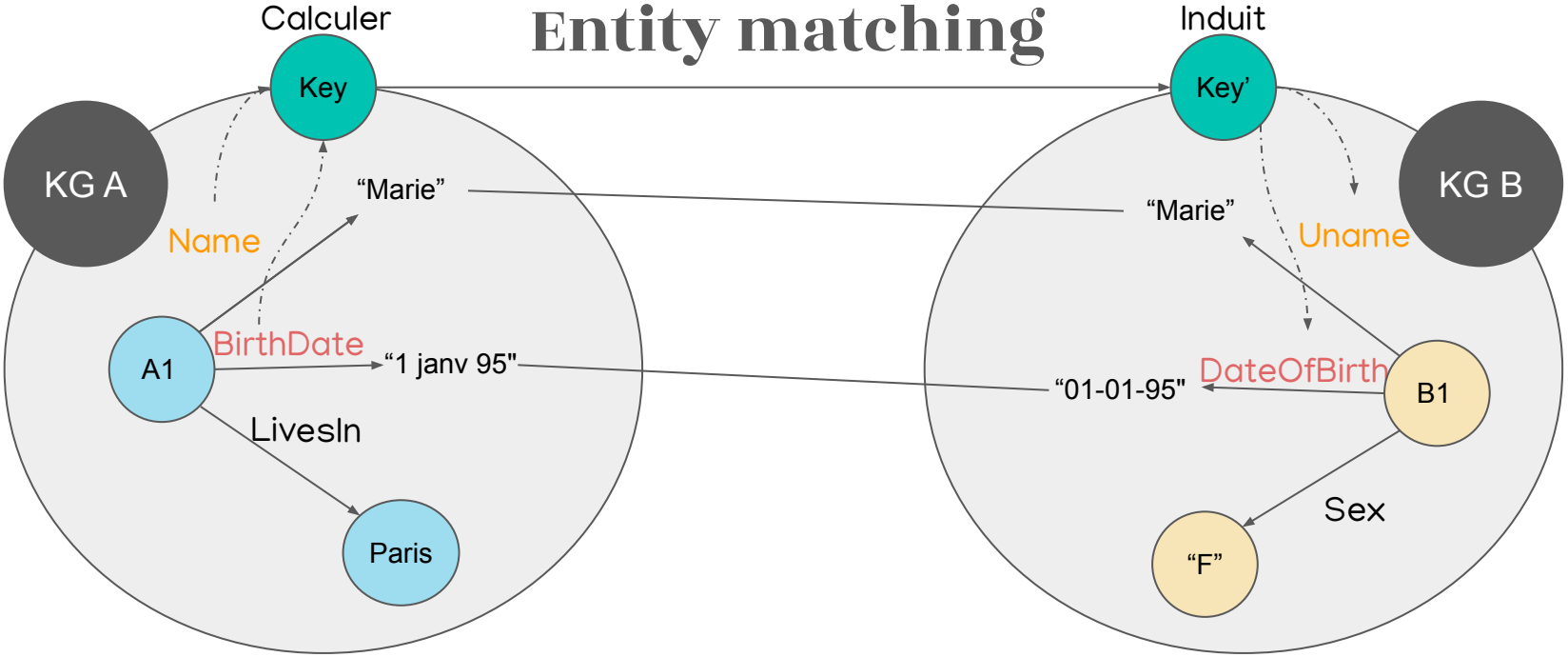
Idée générale
Linking keys



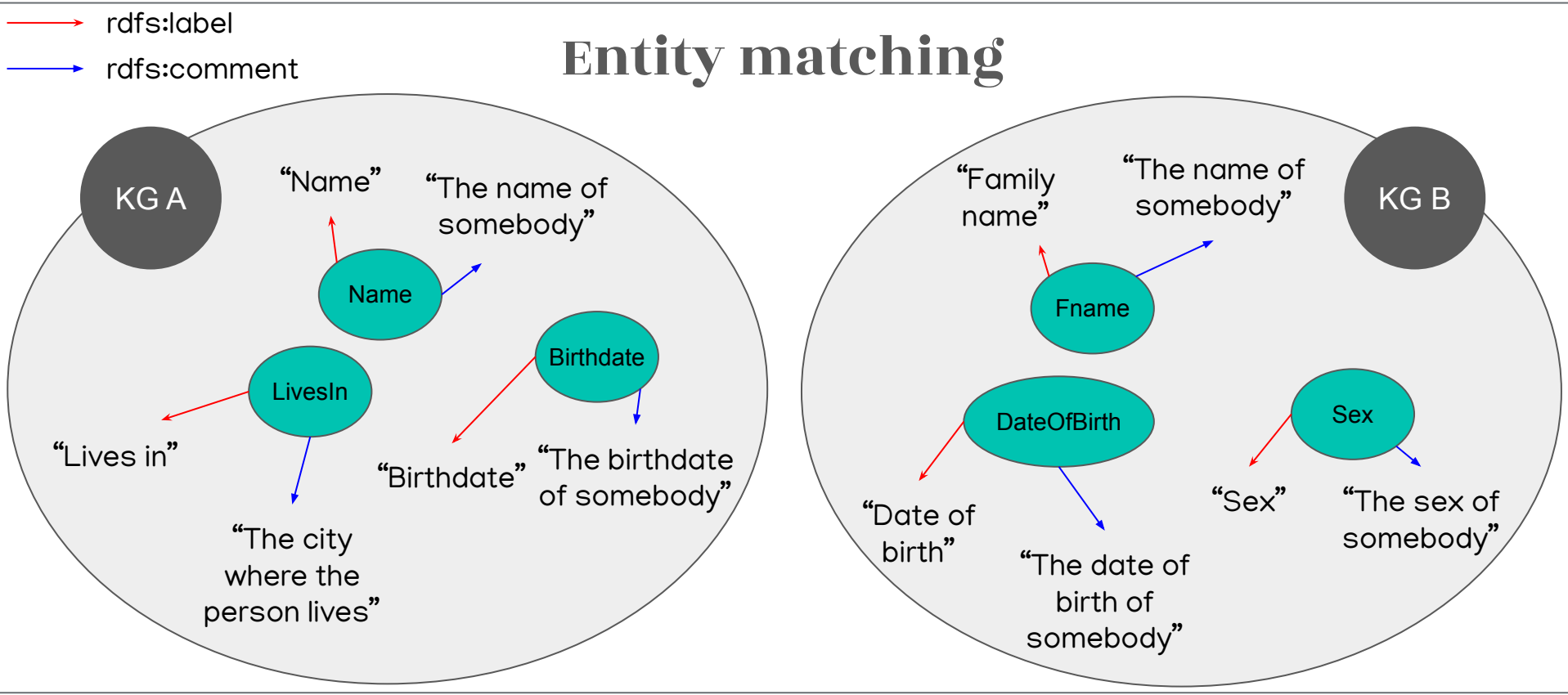
Idée générale
Entity matching



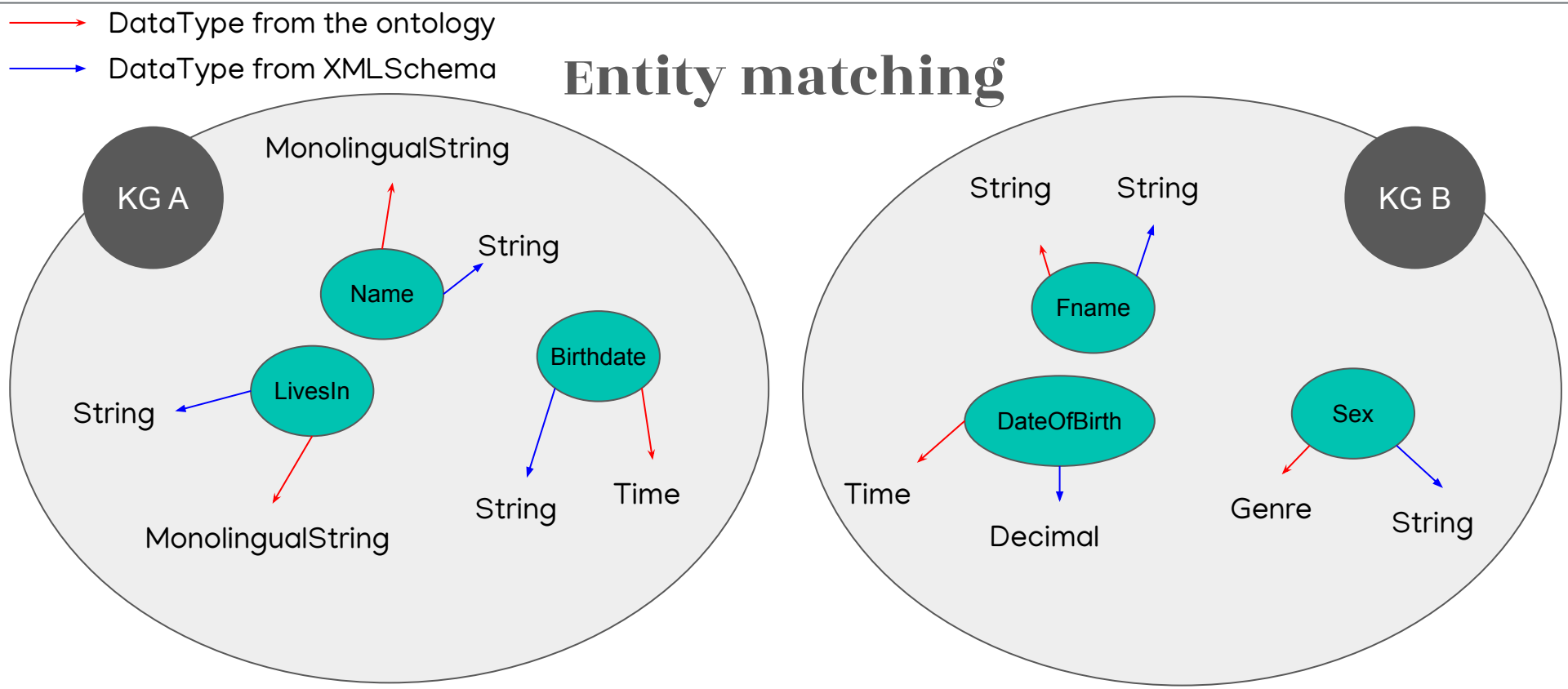
Idée générale
Entity matching



Entity matching



Entity matching



Entity matching

Ontology DataType

Name

Jaro–Winkler

BirthDate

Transform into standard then compare

Population

X ≃ Y ± z%

XML Schema

String

Jaccard index

levenshtein

Decimal

X ≃ Y ± z%